Cognition 157 (2016) 61–76

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Inferring mass in complex scenes by mental simulation

Jessica B. Hamrick [a,*], Peter W. Battaglia [b,c], Thomas L. Griffiths [a], Joshua B. Tenenbaum [c]

[a] Department of Psychology, University of California, Berkeley, United States
[b] Google DeepMind, London, United Kingdom
[c] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States

ARTICLE INFO

Article history:
Received 15 April 2016
Revised 5 August 2016
Accepted 19 August 2016

Keywords:
Mental simulation
Physical reasoning
Probabilistic simulation
Inference
Learning

ABSTRACT

After observing a collision between two boxes, you can immediately tell which is empty and which is full of books based on how the boxes moved. People form rich perceptions about the physical properties of objects from their interactions, an ability that plays a crucial role in learning about the physical world through our experiences. Here, we present three experiments that demonstrate people's capacity to reason about the relative masses of objects in naturalistic 3D scenes. We find that people make accurate inferences, and that they continue to fine-tune their beliefs over time. To explain our results, we propose a cognitive model that combines Bayesian inference with approximate knowledge of Newtonian physics by estimating probabilities from noisy physical simulations. We find that this model accurately predicts judgments from our experiments, suggesting that the same simulation mechanism underlies both peoples' predictions *and* inferences about the physical world around them.

1. Introduction

Consider the scene in Fig. 1a. Despite the difference in size, one can infer that the mass of the forklift is large compared to that of the storage container. Inferences about the physical properties of objects such as mass and friction are critical to how we understand and interact with our surroundings. While they are sometimes specified unambiguously by a small set of perceptible features such as size, material, or tactile sensations, we often access them only indirectly via their physical influence on observable objects. Here, we ask: how do people make such inferences about the unobservable physical attributes of objects from complex scenes and events?

In addition to one-off inferences about properties such as mass, people form beliefs about these properties over time. For example, through experience, people learn that certain materials (e.g., metal) are heavier than others (e.g., plastic). How is it that people learn these attributes? Certainly, people may rely on sensorimotor feedback as they hold and manipulate objects (e.g. Baugh, Kao, Johansson, & Flanagan, 2012). Can people also learn through experience if only visual information about the static and dynamic behavior of such objects is available? If so, what is the mechanism by which they do this?

There is a vast literature on whether (and if so, how) people reason about mass. People are clearly sensitive to mass when reasoning about other physical properties: for example, people's memory for the location of an object is affected by its implied weight (Hubbard, 1997); similarly, people make different judgments about how a tower of blocks will fall down depending on which blocks they think are heavier (Battaglia, Hamrick, & Tenenbaum, 2013). Previous studies of how humans *infer* mass from observed collision dynamics have examined the relative roles of perceptual invariants (Runeson, Juslin, & Olsson, 2000) and heuristics (Gilden & Proffitt, 1994; Todd & Warren, 1982), focusing on judgments about simple one- or two-dimensional (1D or 2D) situations with one or two objects. However, the real world is much more complex: everyday scenes are three-dimensional (3D) and often involve many objects.[1] Moreover, collisions between objects are not the only factor affecting peoples' judgments: for example, there are no collisions in the forklift scene in Fig. 1a, yet we can easily infer what the relative masses of the objects might be.

A question related to *whether* people can make accurate inferences about unobservable physical properties is *how* they make any inferences at all. Sanborn, Mansinghka, and Griffiths (2009, 2013) proposed that inferences could be characterized by a model that performs Bayesian inference over structured knowledge of

* Corresponding author at: 3210 Tolman Hall, University of California, Berkeley, Berkeley, CA 94720, United States.
E-mail address: jhamrick@berkeley.edu (J.B. Hamrick).

[1] We define a 3D scene to be any scene that contains depth information, regardless of whether it is viewed as a 2D projection. We define a 2D scene to be a scene with no depth cues (i.e., it is truly 2D and not a projection of a 3D scene).

http://dx.doi.org/10.1016/j.cognition.2016.08.012
0010-0277/

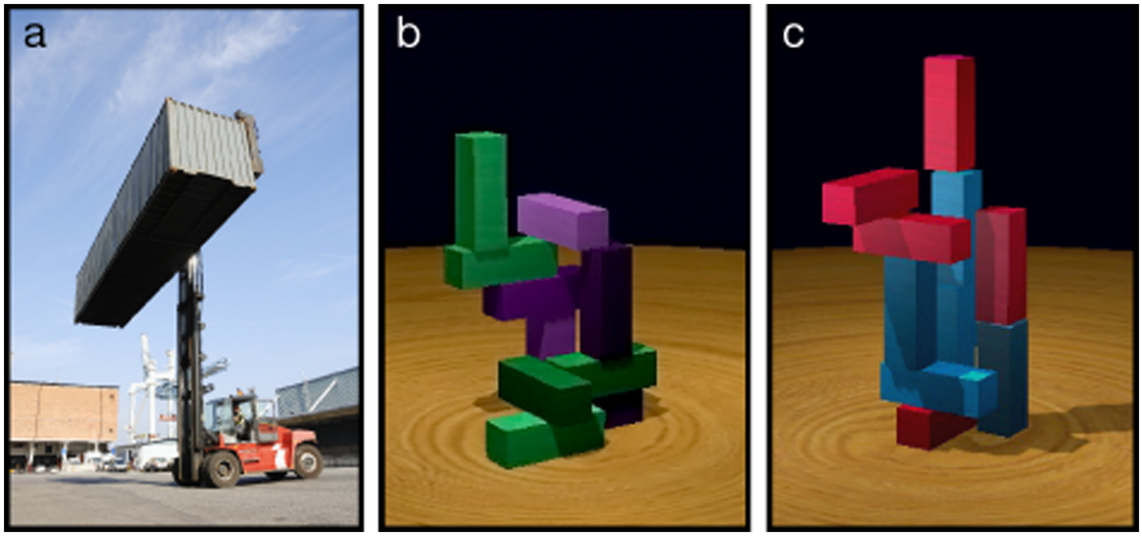

Fig. 1. Three scenes that engage our ability to reason about mass. (a) The forklift's weight counterbalances the container's. (b and c) Two examples of experimental stimuli. If the green blocks in (b) are heavier than the purple blocks, you can predict that the tower will fall down rather than remain standing. If the tower in (c) stays standing, you can infer that the blue blocks are heavier. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Newtonian physics[2] and noisy or uncertain perceptual inputs. In the 2D case, this "Noisy Newton" hypothesis works well for inferring properties like mass because the laws of Newtonian physics (such as conservation of momentum) can be encoded as distributions over random variables such as velocity, where the randomness comes from perceptual uncertainty (Sanborn et al., 2009, 2013; Sanborn, 2014). However, for scenes involving both statics and dynamics, it is not clear where these probabilities should come from. For example, if the forklift in Fig. 1a is about to tip over, you can infer that the storage container is heavier, because if it were not, the forklift would likely remain upright. Where does this "likelihood of remaining upright" come from?

Recent research has proposed that people reason about complex environments using approximate and probabilistic mental simulations of physical dynamics (Battaglia et al., 2013; Hamrick, Battaglia, & Tenenbaum, 2011). They are *approximate* in the sense that they do not analytically solve the exact equations that underly Newtonian physics, but rather estimate the implications of those equations through an iterative process. They are *probabilistic* in that the simulations are non-deterministic, where the stochasticity reflects uncertainty that arises from noisy perceptual processes and imperfect knowledge of the scene. There is a growing body of evidence that people use such approximate and probabilistic mental simulations, including explanations of human judgments of physical causality and prediction in a wide range of scenarios (Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2012, 2014; Hamrick, Smith, Griffiths, & Vul, 2015; Smith & Vul, 2013; Smith, Battaglia, & Vul, 2013; Smith, Dechter, Tenenbaum, & Vul, 2013; Ullman, Stuhlmüller, Goodman, & Tenenbaum, 2014). A similar hypothesis has also been proposed by White (2012), suggesting that simulations are the result of retrieving past perceptual experiences and extrapolating using a forward model.

If people use probabilistic mental simulations to make predictions about physical scenes, then it should be possible for people to use those simulations to estimate the probabilities of different outcomes. These probabilities can be used to make inferences about unobservable physical properties. Indeed, recent work by Ullman et al. (2014) and Gerstenberg et al. (2012, 2014) has provided examples of how simulations might be used in simple 2D scenes to estimate the necessary probabilities for Bayesian inference. However, such an approach has not been applied to the types of complex, 3D scenes that people encounter in the real world.

Using probabilistic simulation to make inferences about unobservable physical properties also suggests a unified framework both for reasoning about individual object-level properties (i.e., that the forklift in Fig. 1a is heavier than the storage container) as well as class- or material-level properties (i.e., that objects made out of stone are heavier than objects made out of plastic). Historically, research has focused on how people make one-shot inferences about the properties of individual objects (Gilden & Proffitt, 1989a; Runeson et al., 2000; Sanborn et al., 2009, 2013; Sanborn, 2014; Todd & Warren, 1982), but not on how these inferences might also play a role in learning class-level properties such as the density of a particular material. We suggest that if Bayesian inference is performed using probabilities obtained through approximate physical simulation, then this could provide an account for *both* one-shot inferences and learning. Specifically, Bayes' rule dictates both how to compute inferences about individual objects, as well as how to integrate multiple pieces of information over time to learn about the properties of classes of objects.

This work is the first to explore people's ability to make inferences about mass in complex scenes that may be either static or dynamic, and addresses two questions regarding this ability. First: *can* people make accurate inferences? To answer this, we present three experiments in which we asked people to make inferences about the relative masses of objects in complex scenes involving both static and dynamic objects. We find that people can form accurate judgments about the relative mass, and that they become increasingly fine-tuned to these properties as they accumulate multiple pieces of information. Second: *how* do people make inferences about properties like mass? We introduce a new cognitive model that uses approximate, probabilistic simulation to estimate probabilities needed by Bayesian inference to produce judgments about the relative mass of objects. When compared to data from our experiments, we find that our model is a good characterization of how people make inferences about the masses of individual objects and how they learn about the mass of a class of objects. Moreover, by replacing the model's simulations with people's own predictions about the future dynamics of the scenes, our

[2] In this context, we take "structured" to mean implicit knowledge of formal physical laws, in contrast to implicit knowledge of naïve physics or explicit knowledge of formal physics. See Section 6for further discussion of how these differing forms of physical knowledge relate.